

Screening insertion libraries for mutations in many genes simultaneously using DNA microarrays

Ramamurthy Mahalingam and Nina Fedoroff*

Biology Department and Biotechnology Institute, 519 Wartik Laboratory, Pennsylvania State University, University Park, PA 16803

Contributed by Nina Fedoroff, April 17, 2001

We describe a method to screen pools of DNA from multiple transposon lines for insertions in many genes simultaneously. We use thermal asymmetric interlaced-PCR, a hemispecific PCR amplification protocol that combines nested, insertion-specific primers with degenerate primers, to amplify DNA flanking the transposons. In reconstruction experiments with previously characterized *Arabidopsis* lines carrying insertions of the maize *Dissociation* (*Ds*) transposon, we show that fluorescently labeled, transposon-flanking fragments overlapping ORFs hybridize to cognate expressed sequence tags (ESTs) on a DNA microarray. We further show that insertions can be detected in DNA pools from as many as 100 plants representing different transposon lines and that all of the tested, transposon-disrupted genes whose flanking fragments can be amplified individually also can be detected when amplified from the pool. The ability of a transposon-flanking fragment to hybridize declines rapidly with decreasing homology to the spotted DNA fragment, so that only ESTs with >90% homology to the transposon-disrupted gene exhibit significant cross-hybridization. Because thermal asymmetric interlaced-PCR fragments tend to be short, use of the present method favors recovery of insertions in and near genes. We apply the technique to screening pools of new *Ds* lines using cDNA microarrays containing ESTs for $\approx 1,000$ stress-induced and -repressed *Arabidopsis* genes.

The rapid expansion of genome sequence information has created a demand for high-throughput techniques in functional genomics. Phenotypic information about gene function often is sought through the analysis of loss- or gain-of-function mutations resulting from DNA insertions (1–16). Transposons, reporter cassettes, gene traps, promoter traps, and *Agrobacterium* T-DNAs all have been used as insertional mutagens in different organisms (1, 2, 5, 6, 17–19). However, the identification of insertion sites remains a rate-limiting methodological challenge in insertional mutagenesis.

DNA sequences flanking insertions have been recovered by plasmid rescue (6) or amplified by several hemispecific PCR methods, such as inverse PCR (20, 21), adapter-ligation PCR (22), vectorette PCR (23), or thermal asymmetric interlaced-PCR (TAIL-PCR) (24, 25). Although laborious and expensive, sequencing of cloned or PCR-amplified flanking fragments unequivocally identifies insertion sites, and databases of insertion-site sequences have been established for mouse embryonic stem cells and *Arabidopsis* (12, 26–27, †). More commonly, insertions are identified by PCR-screening pools of DNA from many insertion lines with gene- and insertion-specific primers (29–33). Variations recently have been described for amplifying and arraying insertion-flanking fragments for screening with gene-specific probes (27, 34). However, all of these methods suffer from the limitation that they permit screening for insertions in only one or a small number of genes at a time (27, 29–35).

The utility of insertional libraries would be enhanced substantially by the ability to screen pools comprising many lines for insertions in many genes at once. In the present work, we describe a PCR-based method of screening pools of DNA for insertions in many genes simultaneously using cDNA microarrays. We have used *Arabidopsis* lines containing maize *Dissociation* (*Ds*) transposons inserted at various locations within and near ORFs to develop the method, and we have investigated its application to detecting insertions in *Arabidopsis* genes induced and repressed by biotic and abiotic stresses.

ation (*Ds*) transposons inserted at various locations within and near ORFs to develop the method, and we have investigated its application to detecting insertions in *Arabidopsis* genes induced and repressed by biotic and abiotic stresses.

Materials and Methods

Transposon Lines. We used 12 previously characterized *Arabidopsis* lines containing maize *Ds* insertions. Transposon-flanking sequences in these lines previously had been analyzed by TAIL-PCR amplification and sequencing,[†] permitting the identification of expressed sequence tags (ESTs) for ORFs at or near the site of insertion. Additional, uncharacterized *Ds* insertion lines were chosen at random. In reconstruction experiments, equal volumes of DNA from plants of different lines were mixed together at a concentration of about 100 ng/ μ l and used as template for TAIL-PCR amplification (24). To construct pools of *Ds* insertion lines with uncharacterized insertion sites for screening, DNA was isolated (Qiagen DNA isolation kit; Qiagen, Chatsworth, CA) from groups of 20 plants, each identified by preliminary screening as containing the transposon and an empty donor site.[†] Equal volumes of DNA from five 20-plant subpools were mixed together to generate 100-plant DNA pools. Approximately 100 ng of DNA was used for TAIL-PCR amplification.

Primers and PCR-Amplification Conditions. The nested, transposon-specific primers were: *Ds*5'1, GCCTTTGCCCTATATGTTTTGCC; *Ds*5'2, GGCATGGCTGGCAATAGCATATTGGC; *Ds*3'2, CCCGACCGGATCGTATCGGTTTTTCGAT; and *Ds*3'3, CGTTTCCGTCCCGCAAGTTAAATATG. The arbitrary degenerate (AD) primers used for TAIL-PCR were: AD7, NTCGASTWTSWGTGTT, and AD8, NGTCGASWGANAWGAA. Additional gene-specific primers were 18–25 mers with a minimum GC content of 50% and average T_m of 60°C.

The thermal profile for PCR amplification of *Arabidopsis* genomic DNA was as follows: initial denaturation at 94°C for 3 min, followed by 30 cycles of 94°C for 30 sec, 60°C for 1 min, 72°C for 2 min, and ending with a final extension at 72°C for 7 min. The PCR mixture (50 μ l) contained 20 ng of DNA, appropriate gene-specific primer pairs (1 μ M each), 1 \times PCR buffer, 2 units of RedTaq (Sigma), and 0.5 mM dNTPs.

The primary TAIL-PCR contained 1 \times PCR buffer supplied by CLONTECH, 0.2 mM dNTPs, 0.5 μ M *Ds*5'1 primer, 2 μ M AD7 primer or 4 μ M AD8 primer, and 1 unit of AdvanTaq (CLONTECH), and the final volume was adjusted to 50 μ l with distilled water. AD primers with less than 128-fold degeneracy were used because they were more efficient than 256-fold degenerate primers (unpublished results). The thermal profile for primary TAIL-PCR was as described in ref. 36. To accom-

Abbreviations: EST, expressed sequence tag; TAIL-PCR, thermal asymmetric interlaced-PCR; *Ds*, *Dissociation*; AD primer, arbitrary degenerate primer.

*To whom reprint requests should be addressed. E-mail: nvf1@psu.edu.

†S. Raina, R.M., F. Chen, and N.F., unpublished data.

The publication costs of this article were defrayed in part by page charge payment. This article must therefore be hereby marked "advertisement" in accordance with 18 U.S.C. §1734 solely to indicate this fact.

moderate the dilution resulting from the pooling of DNA, the original procedure was modified by using 2.5 μ l of the undiluted primary reaction as template for the secondary PCR amplification using the same AD primer and a second, nested *Ds*-specific primer. The secondary PCR consisted of 1 \times PCR buffer, 1 unit of AdvanTaq, the appropriate primers, 0.2 mM dATP, 0.2 mM dTTP, 0.2 mM dGTP, 0.125 mM dCTP, and 0.5 μ l of Cy3-dCTP. The thermal profile for the secondary TAIL-PCR was as described in ref. 36. PCR-amplification products were doubly labeled with 0.5 μ M Cy5-labeled *Ds*5'2 primer, 0.5 μ M gene-specific primer, 1 \times PCR buffer, 1 unit of AdvanTaq, 0.2 mM dATP, 0.2 mM dTTP, 0.2 mM dGTP, 0.125 mM dCTP, and 0.5 μ l of Cy3-dCTP. In some experiments, the fluorescent tags were reversed. The thermal profile for PCR amplification was that described above for the gene-specific amplifications.

For identifying individual *Ds* lines, PCR was carried out with 0.2 μ M EST-homologous primers for each strand, 0.2 μ M *Ds*3'3 or *Ds*5'3 primer, 1 \times PCR buffer, 1 unit of AdvanTaq, and 0.2 mM dNTPs. The template was pooled DNA from 100 *Ds* insertion lines and five subpools of 20 lines each. Plants representing the 20 individual lines comprising the subpool containing the insertion were arranged in five columns and four rows. Leaf samples were pooled from each row and column, and DNA was isolated from these row and column pools to reduce the number of DNA isolations for PCR analyses. The PCR-amplification conditions were as described above.

Microarray Preparation. ESTs selected from an *Arabidopsis* EST collection purchased from the *Arabidopsis* Biological Resource Center (ABRC; Ohio State University) were PCR-amplified from bacterial cultures by using the M13 forward and reverse primers. Fragments of selected genes were amplified from genomic DNA by using gene-specific primers. The amplified ESTs and PCR products were purified by using the Qiagen PCR purification kit. Purified products were evaporated to dryness, and DNA was resuspended in 3 \times SSC (pH 9.0) at a final concentration of 0.5 μ g/ μ l and transferred to a 384-well plate (Costar/Corning). The spots were arrayed in duplicate on silanized slides by using a computer-controlled arraying robot with print pins purchased from Telechem International (Sunnyvale, CA). To detect insertions in many genes simultaneously, we used a cDNA microarray containing \approx 1,200 newly isolated ESTs representing about 1,000 *Arabidopsis* genes induced by biotic and abiotic stresses (R.M., R. Raina, N. Eckardt, and N.F., unpublished results). A detailed protocol for the slide preparation and processing method is available at <http://sgio2.biotech.psu.edu/>.

Hybridization. Excess nucleotides and primers were removed from TAIL-PCR-amplification products (probe) by using a Qiagen PCR purification kit. The probe was evaporated to dryness in a dark-colored Eppendorf tube, resuspended in 7.5 μ l of formamide/3.75 μ l of 20 \times SSC/0.75 μ l of 2% SDS/1.5 μ l of salmon sperm DNA/1.5 μ l of 50 \times Denhardt's solution, denatured in boiling water for 5 min, spun briefly to collect the contents, and applied immediately to a DNA array slide. Slides were covered with a precleaned lifter slip (Erie Scientific Company, Portsmouth, NH) and placed in a hybridization cassette with 15 μ l of distilled water in each chamber to maintain humidity. The cassette was sealed and placed in a water bath at 42°C for 16–18 h. The slides were rinsed at room temperature in 2 \times SSC and 0.2% SDS for 3 min, followed by a 2-min rinse in 1 \times SSC, 1 min in 0.2 \times SSC, and, finally, a 15-sec rinse in 0.05 \times SSC. The slides were spun-dried and scanned immediately in an Axon scanner (Axon Instruments, Foster City, CA).

Sequencing. All ESTs and PCR-amplified fragments on the microarray and PCR-amplified fragments from the confirmatory tests were sequenced using an ABI Prism Dye terminator cycle

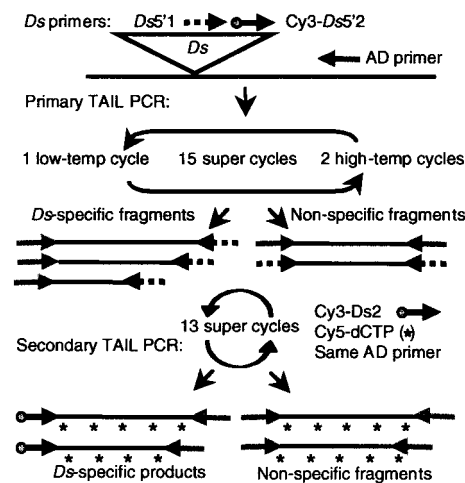


Fig. 1. A diagrammatic representation of TAIL-PCR amplification (25).

sequencing kit (Applied Biosystems) with a T7 primer, a gene-specific primer, or the appropriate *Ds* primer. The sequences were verified using WU-BLAST2 searches from the *Arabidopsis* database (<http://www.arabidopsis.org/search/>).

Results

Using cDNA Microarrays to Detect Insertions in and near Genes: The Principle.

The approach described here is conceptually straightforward: sequences flanking DNA insertions are amplified preferentially by a hemispecific PCR method and hybridized to a cDNA microarray. Flanking fragments that overlap genes represented on the microarray will hybridize to their respective cDNAs, thereby identifying genes containing insertion mutations in or near them. The technical challenges are substantial. First, the method must be readily scalable for use on pools of DNA from many insertion lines. Second, it must be able to detect most or all of the insertions present in the pool. Third, the method must be capable of discriminating among closely related sequences. And fourth, it would be advantageous if the method were able to discriminate between insertions in or very near coding sequences and distant insertions.

To explore these technical challenges, we used previously characterized *Arabidopsis* lines containing maize *Ds* transposon insertions in and near ORFs for which we could identify similar and identical ESTs (5, 37, †). We used the TAIL-PCR procedure developed by Liu and Whittier (25) to amplify sequences adjacent to the transposons because of its simplicity. TAIL-PCR can be applied directly to genomic DNA without additional manipulations, such as the restriction and ligation steps involved in other hemispecific amplification techniques (20, 21). We amplified *Ds*-flanking sequences by using a set of nested, long, transposon-specific primers and a short, arbitrary, degenerate primer that can prime at many sites in the genome (25). The interspersed high and low annealing temperatures, combined with the use of nested, insertion-specific primers in sequential amplification steps, as illustrated in Fig. 1, results in the preferential amplification of the insertion-flanking sequences.

Detecting Insertions by cDNA Microarray Hybridization. To determine whether DNA microarrays could be used to detect insertions in DNA pools, we mixed equal amounts of DNA from three previously characterized *Ds* lines with DNA from additional uncharacterized *Ds* lines to yield pools of 20, 40, 100, 200, and 400 lines. Two lines (201–57, 201–132) had *Ds* insertions in ORFs, and a third (201–16) had a *Ds* insertion 387 bp downstream from the nearest ORF.† Transposon-flanking sequences

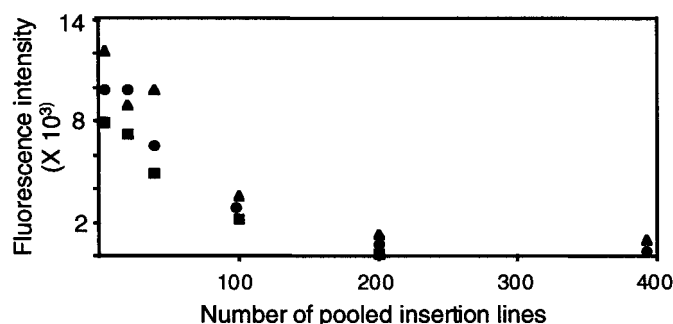


Fig. 2. DNA microarray for detection of insertions. *Ds*-flanking fragments were PCR-amplified in the presence of Cy3-dCTP and template DNA from single plants and pools of different sizes and then hybridized to a DNA array containing ESTs for the disrupted or nearby ORFs. In line 201-16 (▲), the *Ds* is 387 bp downstream of the ORF, whereas those in lines 201-132 (●) and 201-57 (■) are in the ORFs (Table 1).

were TAIL-PCR-amplified in the presence of Cy3-dCTP by using template DNA both from individual lines and from pools. The labeled fragments were hybridized to DNA microarrays containing PCR-amplified ESTs corresponding to the disrupted or nearest ORF in the test lines. Transposon-flanking fragments of all three lines hybridized to the arrayed ESTs (Fig. 2). The fluorescence intensity declined with increasing pool size, but remained easily detectable in pools of 100 or fewer lines. We therefore used 100-line pools in subsequent experiments.

False Positives, False Negatives, and Cross-Hybridization. To assess the specificity of hybridization and the efficiency of identifying insertions by microarray hybridization, we used a larger set of 12 previously characterized *Arabidopsis* lines with *Ds* insertions in and near ORFs for which we could identify similar and identical ESTs.[†] The insertion lines, the position of each insertion, the EST accession number, the homology of the EST to the sequence flanking the insertion, and the degenerate primer (AD) previously used for its amplification[‡] are shown in Table 1. The listed ESTs and 33 additional ESTs with no homology to any of the flanking sequences were PCR-amplified and arrayed on a glass slide (*Materials and Methods*). DNA was isolated from plants containing the insertions, pooled with DNA from additional uncharacterized *Ds* lines, and subjected to TAIL-PCR amplification.

Initially, we used Cy3-dCTP to label the PCR-amplification products internally in a secondary TAIL-PCR. We observed a large number of false positives under these conditions, suggesting cross-hybridization from fragments amplified by the degenerate primers. When a Cy5-labeled, transposon-specific primer was used, fewer false positives were detected, but the signal intensity was substantially lower, because each fragment contains only a single labeled moiety. In the reconstruction experiments described below, we used both Cy3-dCTP and a Cy5-labeled, transposon-specific primer in the secondary TAIL-PCR to take advantage of the higher signal intensity of the internal label while using the terminal label to eliminate false positives. Fig. 3 shows superimposed false-color images obtained by scanning the microarrays at the emission peak of each fluorescent dye. The red and green spots are false positives resulting primarily from hybridization of fragments labeled internally. Yellow represents hybridization of both terminally and internally labeled fragments and is observed for the known *Ds* lines. We expected to detect seven insertions with the *Ds*5' and AD7 primer combination and three with the *Ds*5' and AD8 primer combination. All of the expected insertions were detected (Fig. 3 *a* and *b*). Insertion lines 201-8 and 204-109 were used as

Table 1. Characterized *Ds* insertion lines

	<i>Ds</i> line	Insertion position*	EST	% [†]	Primers [‡]
1	201-16	387 bp dws	219H4T7	100	5', 7
2	201-57	ORF	154I3T7	100	5', 7
			154I3T7	100	3', 7
3	201-132	3' UTR	142P24T7	100	5', 7
			142P24T7	100	3', 7
4	311-39	55 bp dws	163K16T7	100	5', 8
5	201-47	ORF	103K11T7	100	5', 7
6	311-108	239 bp ups	90H7T7	100	5', 8
7	326-8	ORF	98F3T7	100	5', 7
8	201-146	95 bp ups	104E2T7	98	5', 7
9	379-25	ORF	114L10T7	92	5', 7
10	301-7	ORF	PCR frag [§]	100	5', 8
			PCR frag	86	5', 8
11	201-8	ORF	PCR frag	100	3', 4
			164D22T7	70	
12	204-109	129 bp ups	163J17T7	98	5', 2
			86D9T7	71	

UTR, untranslated region; dws, downstream from ORF; ups, upstream of ORF.

*The site of *Ds* insertion was determined by sequencing TAIL-PCR-amplified flanking fragments (28).

[†]Percent nucleotide sequence identity between the EST and the ORF in or near which the *Ds* is inserted.

[‡]*Ds* (5' or 3' end of transposon) primers and AD primers used for TAIL-PCR.

[§]PCR fragments amplified from wild-type genomic DNA by using ORF-specific primers.

negative controls because these lines do not amplify with the primer combinations used. False positives are detected primarily by the hybridization of internally labeled fragments. The intensity of labeling varies considerably from spot to spot. This is not surprising, because both fragment length and the extent of overlap with the ORF vary among TAIL-PCR-amplified fragments. Nonetheless, all of the flanking sequences in the test set of *Ds* insertion lines that previously had been amplified successfully from DNA of individual plants with AD primers 7 and 8 were detected in DNA pools prepared from 100 plants.

The present detection technique must be capable of discriminating among closely related sequences. We assessed the stringency of the hybridization conditions by using two previously characterized *Ds* lines, 326-8 and 301-7, which have insertions in the ORFs of a cytochrome P450 gene (AL161541) and a gene encoding a subtilisin-like serine protease (AB016885), respectively. A fragment flanking each insertion was sequenced and used to identify homologs. Gene-specific primer pairs were used to amplify each of the homologs from genomic DNA for microarray spotting. DNA from *Ds* lines 326-8 and 301-7 was TAIL-PCR-amplified using the *Ds*5' and AD7 primers. The fragments were labeled internally with Cy3-dCTP and used to probe the array. The hybridization intensity declined rapidly with homology for both genes (Fig. 4*a*). Significant cross-hybridization was detected only for sequences that show more than 90% identity with the disrupted gene. These results reveal that cross-hybridization is likely to be a problem only for the very closest homologs.

Identifying Insertions in ORFs. To determine whether the hybridization intensity can provide information about the position of the insertion with respect to the coding sequence, we assessed the hybridization of *Ds*-flanking fragments from lines containing insertions 24, 48, 259, 387, 473, and 576 bp from the nearest ORF. Approximately 500 bp of the corresponding ORFs were amplified and spotted on the array. Terminally and internally labeled *Ds*-flanking fragments amplified by using *Ds*- and gene-

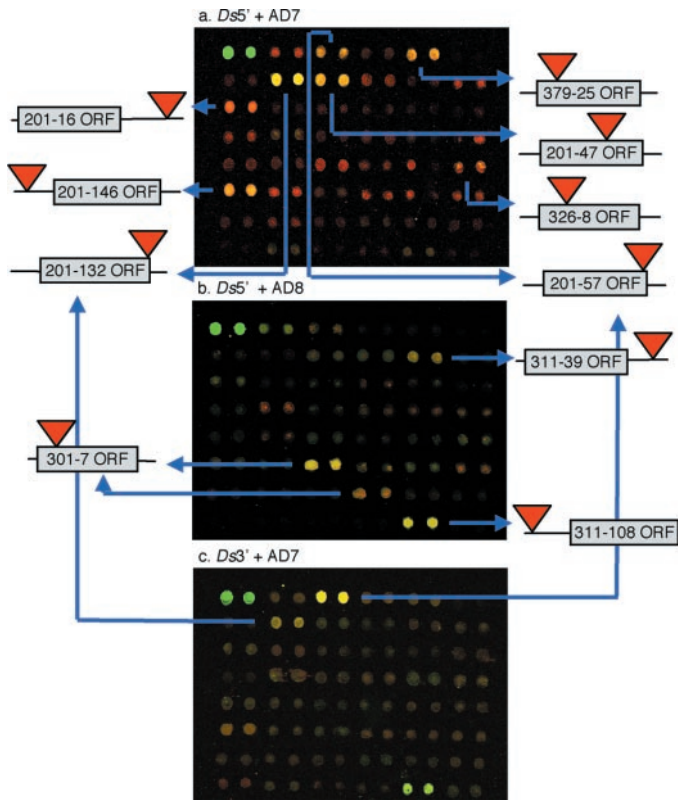


Fig. 3. Detecting all of the insertions in a DNA pool. *Ds*-flanking fragments were TAIL-PCR-amplified from a 100-line DNA pool that included 12 previously characterized *Ds* lines (Table 1) and hybridized to a microarray containing amplified ESTs for the ORFs in or near the known insertion sites, as well as 33 randomly selected ESTs. (a) Fragments labeled with Cy3-*Ds*5'2 primer and Cy5-dCTP. (b) Cy5-*Ds*5'2 primer and Cy3-dCTP. (c) Cy5-*Ds*3'3 primer and Cy3-dCTP. Slides were scanned at the emission peak of each dye, and false-color overlays of the scans are shown. Red and green spots result from hybridization of one label (false positives); yellow spots represent hybridization of both terminally and internally labeled fragments. The pair of green spots at the upper left of each array is a Cy3-labeled, DNA-retention control.

specific primers were pooled and hybridized to the microarray. Because there is a single Cy5-labeled primer per fragment, and the amount of Cy3-dCTP incorporated depends on the length of the fragment, the Cy5/Cy3 ratio should vary with the distance of

the insertion from the ORF. As evident in Fig. 4b, the ratio of terminal to internal label decreases with increasing distance of the insertion from the ORF. Because TAIL-PCR-amplified fragments are heterogeneous in length, they will not exhibit the precise relationship between fragment length and the Cy5/Cy3 ratio observed in the foregoing reconstruction experiment. Nonetheless, Cy5/Cy3 ratios obtained in experiments with TAIL-PCR-amplified fragments flanking ORFs are higher for insertions in ORFs than for insertions at some distance from them. Thus, for the genes with identical ESTs in Table 1, the ratios for insertions in ORFs were between 0.6 and 1.3, whereas the ratios for the two insertions several hundred base pairs away from the respective ORFs were 0.1 or less (Fig. 4b).

Another approach to the identification of insertions within ORFs is the detection of hybridizing fragments extending from both ends of the transposon, because the flanking fragment on only one side of a transposon inserted outside of an ORF has the potential of extending into the ORF. Fig. 3c shows the results of hybridizing TAIL-PCR fragments amplified using primers for the 3' end of *Ds* and primer AD7. Two of the five insertions in ORFs were detected by hybridization of TAIL-PCR fragments extended from both ends of the *Ds* insertion.

Screening Many Genes for Insertions Using DNA Pools. We tested four pools of 100 uncharacterized *Ds* lines for insertions in or near the $\approx 1,000$ genes represented in our 1,200-element array of stress-induced ESTs (see *Materials and Methods*). DNA from one or two lines used in the previous reconstruction experiment was added to each pool, and the corresponding ESTs were added to the array. The four pools were subjected to TAIL-PCR using the *Ds*5' primers and one of the AD primers. The labeled products were hybridized to the 1,200-EST "stress" array. DNA pool 4 amplified with a labeled *Ds*5' primer and the AD7 primer gave a positive signal for a clone encoding the *Zat7* gene, a zinc finger transcription factor (Fig. 5a). The positive control line in this pool is *Ds* 201-16 (Table 1) and contains an insertion 387 bp downstream of a gene encoding a germin-like protein, GLP4 (219H4T7). Strong hybridization was detected for two clones in DNA pool 2 using 5' *Ds* primers and the AD8 degenerate primer (Fig. 5). The sequences of the two clones are identical to the sequence of a glucose transporter gene (*STP1*). The positive control line in this pool, *Ds* 311-39, has a *Ds* insertion 55 bp downstream of a SWI/SNF gene (EST 163K16T7). A few additional ESTs hybridized on both microarrays, albeit less strongly. These failed to amplify with gene- and *Ds*-specific primers, indicating that they resulted from cross-hybridization to

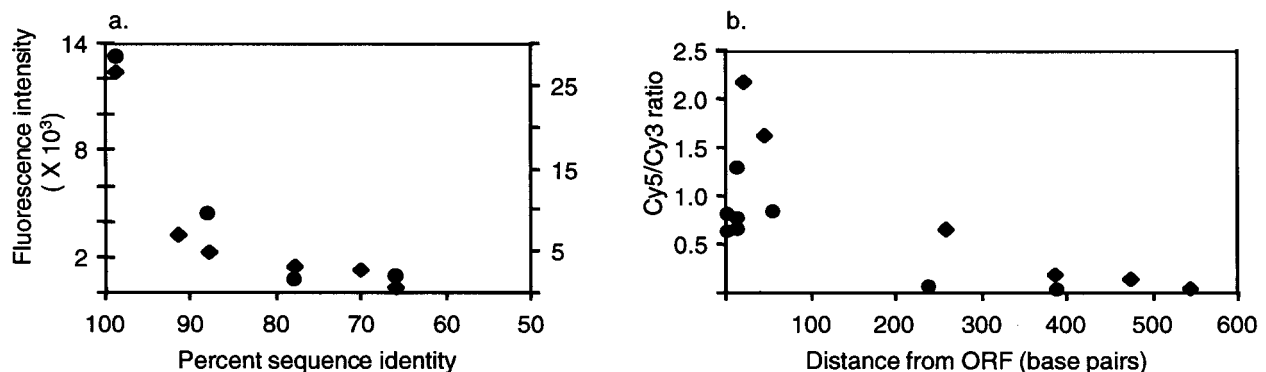


Fig. 4. (a) Cross-hybridization between *Ds*-flanking sequences and homologs. DNA fragments of genes with decreasing homology to the genes disrupted in lines 326-8 (cytochrome p450, \blacklozenge) and 301-7 (subtilisin-type protease, \bullet) were spotted and hybridized to end-labeled PCR fragments amplified from genomic DNA of the two lines. (b) Detecting insertions inside and outside ORFs. *Ds*-flanking fragments were amplified and both terminally and internally labeled from lines with insertions 24, 48, 259, 387, 473, and 576 bp from the ORF by using *Ds*- and gene-specific primers (\blacklozenge) or *Ds*-specific and degenerate primers under TAIL-PCR conditions (\bullet).

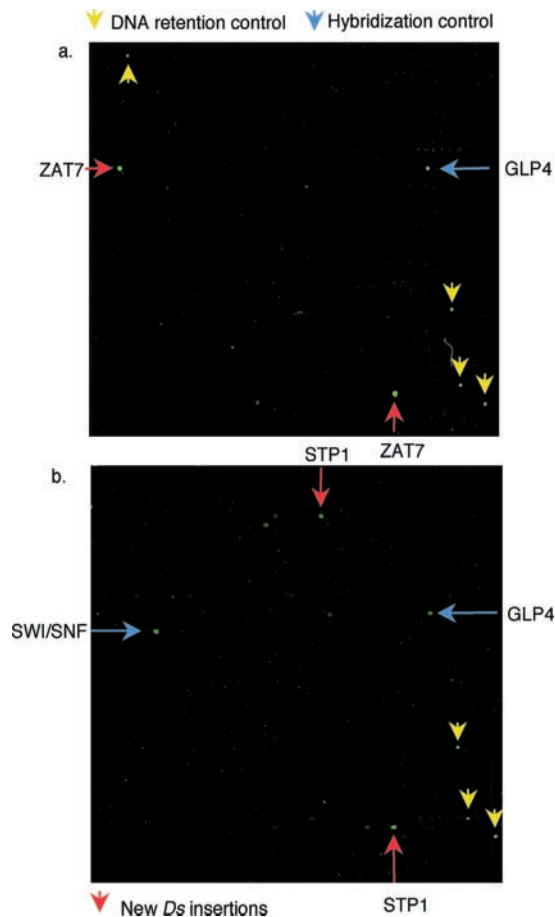


Fig. 5. Screening for insertions in many different genes. DNA pools prepared from 100 uncharacterized *Ds* lines were TAIL-PCR-amplified in the presence of Cy3-*Ds*5'2 primer and degenerate primer AD7 (a) or AD8 (b) and hybridized to a 1,200-EST *Arabidopsis* "stress" gene microarray (see *Materials and Methods*).

sequences with low homology to *Ds*-flanking fragments in the pool.

Identifying Individual Insertions. To confirm the presence of an insertion in a gene and determine its location, we used gene-specific primers complementary to both strands of the cDNA encoding the gene believed to be disrupted, together with the *Ds* primers. Fragments were amplified by using template DNA from the 100-line pool and five subpools comprising DNA from 20 lines. As shown in Fig. 6 for the *Ds* insertion in the STP1 gene (Fig. 5b), fragments of the same size were amplified from the 100-line pool and one of the 20-line subpools. The individual plant containing the insertion was identified by using further column and row pools prepared from individual plants in the 20-line subpool. The identity of the amplified fragment was established unequivocally by reamplifying and sequencing the fragment, revealing that *Ds* had inserted into an exon of the STP1 gene. By the same procedure, the insertion identified by hybridization of a *Ds*-flanking fragment to the *Zat7* clone was found to have a *Ds* insertion 77 bp downstream of the *Zat12* gene, a close homolog of *Zat7* with regions of nucleotide sequence identity exceeding 95%.

Discussion

We have described a method that can be used to detect DNA insertions in many genes simultaneously. We have modified

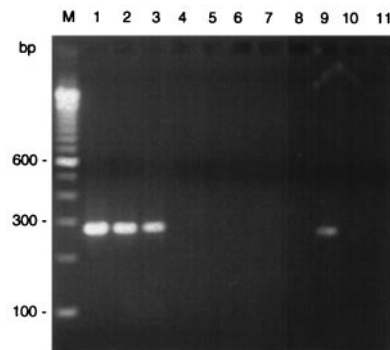


Fig. 6. Identifying the individual *Ds* line. DNA fragments were PCR-amplified by using EST- and *Ds*-specific primers from DNA extracted from the 100-line pool (lane 1), 20-line pool containing the insertion (lane 2), column pools (lanes 3–7), and row pools (lanes 8–11) of leaves from individual plants. The amplified fragments were separated on an agarose gel.

TAIL-PCR (24, 38), a simple, hemispecific PCR-amplification method, for the amplification of DNA pools extracted from sets of many individual *Arabidopsis* lines with *Ds* transposon insertions. We showed that under the conditions used, microarray hybridization of TAIL-PCR-amplified fragments can detect individual *Ds* insertion lines in a DNA pool comprising as many as 100 lines. Moreover, we showed that all of the transposon-flanking sequences that can be identified by a given transposon-specific and degenerate primer pair using DNA from a single line also can be detected using a 100-line DNA pool as template. Because 80–90% of flanking sequences can be amplified using the AD7 and AD8 degenerate primers in combination with a nested primer set for the 5' end of the *Ds* transposon (unpublished data), combining TAIL-PCR amplification with microarray hybridization has the potential to detect most of the gene-proximal insertions in a large pool of DNAs.

Not unexpectedly, the hybridization intensity detected by microarray hybridization varies substantially among the insertion lines tested. Disparities can arise by differential amplification or labeling of fragments, but they also may be caused by cross-hybridization of related sequences. To systematically analyze the effect of fragment length and homology and to identify false positives generated by spurious amplification of fragments, we used two different fluorophores to label amplified fragments both terminally and internally. Both the specific and the degenerate primers can amplify fragments during the low-temperature amplification cycles of TAIL-PCR (24, 38). But, because only the degenerate primer is the same in both the primary and secondary amplifications, it is likely to make the greatest contribution to generating nonspecific fragments. Consonant with this reasoning, we observed substantially more spurious hybridization of internally than of the terminally labeled fragments. By contrast, both internally and terminally labeled fragments hybridize to the ESTs at or near *Ds* insertion sites.

The use of both terminal and internal labels also provides information about the position of the insertion relative to the nearest ORF. The greater the distance of the insertion from the gene, the longer the fragment must be to hybridize to the cDNA on the microarray. This is reflected in the decreasing ratio of terminal to the internal label observed with increasing distance of the insertion site from the ORF (Fig. 3b). Because TAIL-PCR-amplified fragments tend to be short (0.4–1 kb), the present method inherently favors detection of insertions in and very near ORFs, where they are most likely to disrupt gene function. Fragments flanking insertions within exons will hybridize to the corresponding cDNA even if short. Moreover,

because *Arabidopsis* introns are generally short (39), even fragments flanking intron insertions are likely to hybridize.

In principle, it should be possible to identify insertions within ORFs uniquely by the ability of insertion-flanking fragments extending from both ends of the *Ds* to hybridize to the same EST. In practice, the combinations of *Ds*3' and AD primers we have used so far amplify fewer *Ds*-specific fragments than *Ds*5' and AD primer combinations (Fig. 3). This may be improved by further optimizing the *Ds*3' primers. However, the observation that *Ds* insertions cluster around the translational start site (26, †), combined with the fact that TAIL-PCR-amplified fragments tend to be short, already favors recovery of *Ds* insertions either in ORFs or sufficiently near them to affect gene activity.

We screened four pools of 100 previously uncharacterized *Ds* insertion lines and identified two new insertions by using a set of 1,200 short ESTs representing about 1,000 stress-induced and -repressed *Arabidopsis* genes. As we have reported previously, some of the *Arabidopsis* lines in which *Ds* excision has occurred do not contain a reinserted element, either by virtue of its genetic segregation from the empty donor site or because the transposon failed to reinsert (37, †). Because such lines were not eliminated from the collection used here, we estimate that the 400 lines represent ≈200 new insertions. In addition, almost all of the clones on the stress microarray represent *Rsa*I-restricted cDNA fragments ranging in size from 100–500 bp. Such cDNA fragments are likely to detect only some of the insertions in and near the corresponding genes. Moreover, as noted above, the present method is likely to favor identification of insertions in or very near genes, and a single *Ds*5' and AD primer combination usually amplifies 60–70% of transposon-flanking sequences (unpublished data). Considering that *Arabidopsis* contains

≈26,000 genes (28), the recovery of two new insertions in the present subset of genes is roughly what might be expected. Additional improvements in detecting all insertions present in a given pool may be achieved by optimizing primer combinations, pooling fragments TAIL-PCR-amplified with different primer combinations, and using longer cDNAs.

The method we have devised is generalizable to any DNA insertional mutagen in any organism and is significantly more efficient than current methods, all of which involve repeated gene-by-gene screening of either pooled DNAs from insertion lines or PCR-amplified, insertion-flanking fragments (20–35). It can be used as easily with large microarrays containing the entire complement of ESTs available for a given organism as with smaller customized microarrays containing cDNAs from a particular organ or tissue, activated by a particular environmental stimulus, or encoding enzymes in a restricted set of metabolic pathways. The number of insertion lines that can be screened simultaneously can be increased further over the number used here either by improving the TAIL-PCR-amplification protocol or by the simple expedient of combining labeled fragments amplified from several pools for simultaneous hybridization. The elegance of the strategy lies in the fact that the screening need be done only once with each DNA pool to identify all of the insertions. A database of such hybridization results would make it possible for multiple investigators to identify pools containing insertions in genes of interest without additional screening.

We thank Drs. Surabhi Raina, Nancy Eckardt, and Ramesh Raina for cDNA clones and *Ds* lines and John Szot for assistance in microarray preparation. This work was supported by National Science Foundation Plant Genome Grant 9872629.

- Amsterdam, A., Yoon, C., Allende, M., Becker, T., Kawakami, K., Burgess, S., Gaiano, N. & Hopkins, N. (1997) *Cold Spring Harbor Symp. Quant. Biol.* **62**, 437–450.
- Amsterdam, A., Burgess, S., Golling, G., Chen, W., Sun, Z., Townsend, K., Farrington, S., Haldi, M. & Hopkins, N. (1999) *Genes Dev.* **13**, 2713–2724.
- Azpiroz-Leehan, R. & Feldmann, K. A. (1997) *Trends Genet.* **13**, 152–156.
- Cooley, L., Kelley, R. & Spradling, A. (1988) *Science* **239**, 1121–1128.
- Fedoroff, N. V. & Smith, D. L. (1993) *Plant J.* **3**, 273–289.
- Hamilton, B. A., Palazzolo, M. J., Chang, J. H., VijayRaghavan, K., Mayeda, C. A., Whitney, M. A. & Meyerowitz, E. M. (1991) *Proc. Natl. Acad. Sci. USA* **88**, 2731–2735.
- Hehl, R. (1994) *Trends Genet.* **10**, 385–386.
- Bechtold, N. & Pelletier, G. (1998) *Methods Mol. Biol.* **82**, 259–266.
- Martienssen, R. A. (1998) *Proc. Natl. Acad. Sci. USA* **95**, 2021–2026.
- Ross-Macdonald, P., Coelho, P. S., Roemer, T., Agarwal, S., Kumar, A., Jansen, R., Cheung, K. H., Sheehan, A., Symoniat, D., Umansky, L., et al. (1999) *Nature (London)* **402**, 413–418.
- Walbot, V. (1992) *Annu. Rev. Plant Physiol.* **43**, 49–82.
- Wiles, M. V., Vauti, F., Otte, J., Fuchtbauer, E. M., Ruiz, P., Fuchtbauer, A., Arnold, H. H., Lehrach, H., Metz, T., von Melchner, H., et al. (2000) *Nat. Genet.* **24**, 13–14.
- Zwaal, R. R., Broeks, A., van Meurs, J., Groenen, J. T. & Plasterk, R. H. (1993) *Proc. Natl. Acad. Sci. USA* **90**, 7431–7435.
- Speulman, E., Metz, P. L., van Arkel, G., te Lintel Hekkert, B., Stiekema, W. J. & Pereira, A. (1999) *Plant Cell* **11**, 1853–1866.
- Meissner, R. C., Jin, H., Cominelli, E., Denekamp, M., Fuertes, A., Greco, R., Kranz, H. D., Penfield, S., Petroni, K., Urzainqui, A., et al. (1999) *Plant Cell* **11**, 1827–1840.
- Chin, H. G., Choe, M. S., Lee, S. H., Park, S. H., Koo, J. C., Kim, N. Y., Lee, J. J., Oh, B. G., Yi, G. H., Kim, S. C., et al. (1999) *Plant J.* **19**, 615–623.
- Dougherty, B. A. & Smith, H. O. (1999) *Microbiology* **145**, 401–409.
- Cecconi, F. & Meyer, B. I. (2000) *FEBS Lett.* **480**, 63–71.
- Roos, D. S., Sullivan, W. J., Striepen, B., Bohne, W. & Donald, R. G. (1997) *Methods* **13**, 112–122.
- Ochman, H., Ajioka, J. W., Garza, D. & Hartl, D. (1989) in *PCR Technology: Principles and Applications for DNA Amplifications*, ed. Erlich, H. A. (Stockton, New York), pp. 105–111.
- Dalby, B., Pereira, A. J. & Goldstein, L. S. (1995) *Genetics* **139**, 757–766.
- Lagerstrom, M., Parik, J., Malmgren, H., Stewart, J., Pettersson, U. & Landegren, U. (1991) *PCR Methods Appl.* **1**, 111–119.
- Eggert, H., Bergemann, K. & Saumweber, H. (1998) *Genetics* **149**, 1427–1434.
- Liu, Y. G., Mitsukawa, N., Oosumi, T. & Whittier, R. F. (1995) *Plant J.* **8**, 457–463.
- Liu, Y. G. & Whittier, R. F. (1995) *Genomics* **25**, 674–681.
- Parinov, S., Sevugan, M., De, Y., Yang, W. C., Kumaran, M. & Sundaresan, V. (1999) *Plant Cell* **11**, 2263–2270.
- Tissier, A. F., Marillonnet, S., Klimyuk, V., Patel, K., Torres, M. A., Murphy, G. & Jones, J. D. (1999) *Plant Cell* **11**, 1841–1852.
- Walbot, V. (2000) *Nature (London)* **408**, 794–795.
- Ballinger, D. G. & Benzer, S. (1989) *Proc. Natl. Acad. Sci. USA* **86**, 9402–9406.
- McKinney, E. C., Ali, N., Traut, A., Feldmann, K. A., Belostotsky, D. A., McDowell, J. M. & Meagher, R. B. (1995) *Plant J.* **8**, 613–622.
- Krysan, P. J., Young, J. C., Tax, F. & Sussman, M. R. (1996) *Proc. Natl. Acad. Sci. USA* **93**, 8145–8150.
- Krysan, P. J., Young, J. C. & Sussman, M. R. (1999) *Plant Cell* **11**, 2283–2290.
- Winkler, R. G. & Feldmann, K. A. (1998) *Methods Mol. Biol.* **82**, 129–136.
- Yephremov, A. & Saedler, H. (2000) *Plant J.* **21**, 495–505.
- Young, J. C., Krysan, P. J. & Sussman, M. R. (2001) *Plant Physiol.* **125**, 513–518.
- Tsugeki, R., Kochieva, E. Z. & Fedoroff, N. V. (1996) *Plant J.* **10**, 479–489.
- Smith, D., Yanai, Y., Liu, Y. G., Ishiguro, S., Okada, K., Shibata, D., Whittier, R. F. & Fedoroff, N. V. (1996) *Plant J.* **10**, 721–732.
- Hui, E. K., Wang, P. C. & Lo, S. J. (1998) *Cell. Mol. Life Sci.* **54**, 1403–1411.
- Arabidopsis* Genome Initiative (2000) *Nature (London)* **408**, 796–815.